PSYCHOLOGICAL STUDIES IN SEMANTICS: II. RELIABILITY OF FREE ASSOCIATION TESTS*

Department of Psychology, Dartmouth College

THEODORE F. KARWOSKI AND FRED BERTHOLD, JR.

A. INTRODUCTION

The free association technique offers a method for studying relations between words. It has been widely used for various purposes, but the results are often difficult to appraise because investigators use different systems of classification for free association responses. Moreover, it has been claimed that free association to words implies a special set on the part of the subject which is unlike the attitude that the subject would take to things as experienced in every day life. In responding freely to words the subject is said to act as if he were playing with words and their connections. In the initial article of this series free responses to words, objects, and drawings were presented with results which indicate that the attitudes involved in all three situations are essentially similar (4). Finally, there is the question of the reliability of the free association test. Reliability, in its broadest meaning, of any classification system for free association responses involves three problems: (*a*) the consistency of the subject responses with repetition of the test, (*b*) the degree to which different judges agree in classifying the same responses, and (*c*) the degree to which each category in the system is independent of the other categories. The free association technique requires standardization so that the results from it can be properly interpreted.

In this article a system of classification of the relations between words is presented. We believe this system offers possibilities for standardization. About 90 per cent of the free association responses can be classified within six categories. Although the categories are logical in form, they were developed in the effort to get at the psychological factors underlying free association responses. A completely psychological analysis cannot be made without introspections for each response. Our classification was envisaged from the point of view of deriving a logical classification which would make sense psychologically also. Therefore, our system of classifying is a *psycho-logical* one. This article is intended to show that the logical and psychological are sufficiently related to warrant the use of a logical system of classifying to get

---

at the psychological intention of the subject. The reason for using a logical system is twofold: (*a*) it is possible to apply a more strict and unvarying definition to each category thereby reducing one factor which would make for lack of objectivity between different judges, and (*b*) fewer categories are needed.

The empirical connection between logic and psychology in our classification is based on previous studies which indicate that thinking is in terms of identities and opposites (5, 6, 8). With the assistance of Professor Francis Gramlich, a logician, a set of categories was constructed indicating degrees of identification and contrast. It is interesting to note that Plato (9) in the *Phaedo* discussed various kinds of similarities and contrasts which strongly resemble our own categories.[1]

B. THE CATEGORIES

The following categories are the same as those presented in the initial article except that here the definitions are in more simple English. The *operational* and *phrase completion* categories are added because they seem psychologically significant even though we will not be directly concerned with them in this article.

1. *Essential similarity.* Ex.—large–big, rough–rugged.

The sense of the response word is the same as the sense of the stimulus word; they have the same meaning.

2. *General Identification.* Ex.—cabbage–vegetable, hand–arm, dark–color, heavy–weight, kill–destroy.

The response makes a generalization about the stimulus word. The meaning of the stimulus word is always included in the meaning of the response word. But the meaning of the response word always contains other meanings than that of the stimulus word.

3. *Specific Identification.* Ex.—ocean–Pacific, friend–Tom, kill—murder, round–sphere, quick–flash.

The response picks out a specific meaning which is included in the stimulus. The meaning of the stimulus word always contains other meanings than that of the response word. But the meaning of the response word is always included in the meaning of the stimulus word.

4. *Contingent Identification.* Ex.—egg–breakfast, music–roommate, house–comfort, mother–good, face–beard.

The stimulus and the response are not *logically* similar. But because

---

[1]His example of essential contrast was "odd" and "even." "3" is a *specific* kind of "odd" while "odd" is the *general* attribute of "3." While "2" and "3" are not necessarily opposites, they are *contingently opposed* by virtue of the fact that "2" always implies evenness while "3" always implies oddness.

of everyday experience, the meaning of the stimulus word and the meaning of the response word have come to be associated as characteristic of the same object or act. The things which are related as having common elements due to the social usage fall in this class.

      5. *Essential Opposition.* Ex.—hot–cold, fill–empty.

The sense of the stimulus word is directly opposed to the sense of the response.

      6. *Contingent Opposition.* Ex.—foot-hand, house-barn, bread-water, ball–bat, flower–bee.

The stimulus and the response are not *logically* opposed. But because of everyday experience, the meaning of the stimulus word and the meaning of the response word have come to be associated as opposed to each other. The oppositions which have been built up by social usage and pressure fall in this class.

      7. *Operational Definition.* Ex.—hit-ball, floor-wash, food-taste, kill–man, hill–ski.

By operation is meant a response which implies action upon an object or movement to a new position. When a verb is given in response to a verb it does not necessarily imply action. It may be merely defining the stimulus word. The response examples above are also instances of contingent identification.

      8. *Phrase Completion.* Ex.—boy–friend, morning–good, fill–harmonic, take–away, hair–cut.

By phrase completion we mean automatized word relations from everyday life which although logical have become habit.

The first six categories constitute the basic logical categories. Of these the first four concern degrees of identification, the next two concern implications of opposition or contrast. After being classified in terms of these six categories the response words may be reclassified in terms of non-logical concepts. Operations and phrase completions are non-logical. These concepts have been included in the list of definitions, because they occur frequently in free association responses. Therefore, they yield a sufficient measure to work with. They may be measures of subcritical logic. Operationalism may be an indication of the logic of action (1) and phrase completion the logic of non-critical if not sub-conscious thinking. However, words so classified were also classified in one of the first six categories. Other non-logical categories could be added for the convenience of the particular investigator such as personal response, value judgments, clangs, synaesthetic analogies, and imagery. Thus a twofold or double barrelled classification of *psychological* and purely psychological categories may be derived to serve both general and clinical studies of free association. We have actually ex-

amined our data for value judgments, emotional responses, clangs, operationalism, and phrase completion. We found but few responses in these categories except in operationalism and phrase completion. Some adjectives elicit synaesthetic associations, and a few nouns elicit value judgments. All concepts except operationalism have been omitted from this analysis to focus attention on the main *psycho-logical* categories.

C. Agreement between Judges

A test of 59 words was given to groups of college students. It was repeated 40 days later to determine reliability. Two advanced students of psychology and philosophy were given the definitions of the categories and told to classify the responses in terms of these definitions.

It happened that 20 of the stimulus words in this list were the same that the senior author and Professor Gramlich gave to college students a year before and classified in the same categories. Therefore, it was possible to study the amount of disagreement between the two sets of judges and to see what the nature of the disagreement was. In the following discussion K-G refers to judges Karwoski and Gramlich; B-K refers to judges Berthold and Kuhn. The 20 words were: reach, take, receive, sweet, young, glad, still, friend, story, hair, make, most, every, come, kill, deep, few, dark, large, and early.

Table 1 is an analysis of the judgments of K-G and B-K. In the first column the categories are indicated, followed by columns for number and frequency of responses; the next four columns deal with the disagreement between judges. The last four columns indicate the number of words still disagreed on after discussion between the two sets of judges. In this discussion Professor Gramlich was not present because of Naval duties, but the senior author represented both K-G judges. Operational responses were also classified in one of the other six categories; therefore, they do not appear in the totals. There were no disagreements in classifying operational responses. The data in the table are based on responses from 200 students—100 by K-G; 100 by B-K.[2]

The results show that judges B-K disagree with judges K-G on 31.9 per cent of the words classified. Judges K-G disagree with judges B-K 32.3 percent. When the frequencies of the responses were taken into account the disagreements were 17.5 per cent and 21.2 per cent respectively. The de-

[2]Since we were mainly interested in agreement between judges on the six basic logical categories, responses which were classified as phrase completion, reiteration or clang by either set of judges were dropped from the analysis.

TABLE 1
AGREEMENT BETWEEN JUDGES

	Category	No. of responses in category	No. of disagreements	% of disagreement	Frequency of responses in category	Disagreement based on frequency	% of disagreement based on frequency	No. of disagreements after conference	% disagreement after conference	Disagreement after conf. based on frequency	% of disagreement after conf. based on frequency
Classified by B and K — K and G disagree	1	29	3	.103	480	3	.006	2	.069	2	.004
	2	35	10	.286	366	113	.309	8	.229	84	.230
	3	6	3	.500	37	10	.270	2	.333	8	.216
	4	25	11	.440	137	47	.343	8	.320	36	.263
	5	13	9	.692	36	31	.861	8	.615	30	.833
	6	59	18	.305	192	60	.313	11	.186	37	.193
	*7	13	0	.000	31	0	.000	0	.000	0	.000
	tot	167	54	.323	1248	264	.212	39	.234	197	.158
Classified by K and G — B and K disagree	1	27	1	.037	766	3	.004	1	.002	3	.004
	2	32	7	.219	225	25	.111	5	.119	21	.093
	3	17	14	.824	158	129	.816	11	.262	63	.400
	4	29	15	.517	107	50	.467	13	.310	24	.224
	5	12	8	.667	28	18	.643	4	.333	14	.500
	6	49	8	.163	143	25	.175	8	.190	25	.175
	*7	13	0	.000	32	0	.000	0	.000	0	.000
	tot	166	53	.319	1427	250	.175	42	.253	150	.105

*Responses in Category 7 do not appear in totals since they were also classified in one of the logical categories.

crease of disagreement with frequencies indicates that there is less tendency to disagree on the more common responses.

Our sample is too small to make profitable a minute analysis of the data. There is less disagreement in the categories for essential contrast, essential identification, and contingent identification than for general identification, contingent contrast, and specific identification. Of the latter three categories, specific identification is easiest to classify.

The exercise of the two sets of judges discussing their disagreements has uncovered a basic problem involved in classifying free association responses. Judges K-G did not adhere strictly to the logical definitions when they felt they knew the psychological intention of the subject. Judges B-K on the other hand, completely disregarded the psychological intention of the subject and maintained a strictly logical standard. It, therefore, turns out that this experiment—in so far as it deals with agreement between judges—was not strictly controlled. The judges represented two different frames of reference. Therefore, the disagreements indicated in this article are probably considerably greater than would be the case if the frame of reference between the two sets of judges was identical. It is our belief that the amount of agreement found in this analysis (Table 1) is an indication of the degree to which the logical and psychological interpretations of word meaning overlap.

The table shows that 5 per cent of the disagreements were resolved when the two sets of judges met. This was not due to a shift in the frame of reference of either set of judges. It was due to the fact that some words have competing levels of meaning. For example, the word *force* may equally well be taken as a verb meaning to coerce or as a noun meaning strength. The word *present* is similar. Another kind of word with competing levels of meanings is exemplified by *fair*, which may mean just or light in color or beautiful. Even a common relationship like men-boy may be regarded from different levels of meaning; on the basis of age it is a contrasting relation. Usually in such cases little can be done since one interpretation of the response is logically just as valid as another. In some cases, however, the individual judges may have failed to see a level of meaning that is brought out when all the judges get together. If this level of meaning seems more probable in representing the subject's intention and yet does not violate the logical classification, the judges were able to resolve the disagreement and reclassified the response in the logical category demanded by the meaning accepted by the judges. Such changes account for the 5 per cent decrease in disagreement.

The disagreement on the category, general identification, were largely a matter of frame of reference. For instance, 7 out of the 14 disagreements came from the stimulus word reach. Such responses as take, get, stretch, arm, hand, K-G considered as including the idea of reach, and therefore classified these responses as general identification. Judges B-K regarded these responses as being merely contingently associated. Logically judges B-K have the stronger position but in terms of common parlance judges

K-G believed that *to reach* is implied by to get, to touch, to stretch and implies arm, hand. Although judges B-K were also aware of the implications of common parlance, they stood by the strict logical definitions of the categories. This is a clear cut example of the operation of two frames of reference—the logical and the *psycho-logical.*

The disagreement on the category, specific identification, can be traced to a similar clash in frame of reference. For instance, for the stimulus word *take* the response steal, seize, or grab are logically special forms of taking; however, psychologically they may be considered essentially coercive and therefore in the category, essential similarity.

Due to our small sample we did not find enough responses in the category, contingent contrast, to uncover any predominant factor causing disagreement. Introspection of judges shows that this category is the hardest to use and the statistical evidence also supports the idea that it is the least reliable of the categories.

Besides *psycho-logical* factors contributing to disagreement between judges the nature of the stimulus word is important. Two of the 20 words, reach and take, produced 19 or 35 per cent of all the disagreements. It was pointed out above that for these stimuli the responses were often conflicting rather than over-lapping in their logical and psychological implications.

In summary of the facts of agreement between judges it would seem that the order of disagreement is about 33 per cent. Part of this figure is due to differences in attitude that the two sets of judges took in applying their definitions of the categories. Another cause of difference may be attributed to the stimulus words, some give responses which cause difficulty unless a definite frame of reference is consistently used by the judges. Other disagreements involve levels of meaning in which the responses are interpreted by the judges and the degree of goodness of either the logical or psychological implications for a particular category operate with all stimulus words and are probably not subject to control by the judges. In some measure the frame of reference and the choice of stimulus words can be controlled and probably agreement of the order of 80 per cent between judges is not unreasonable to expect. Looking at the figure of 33 per cent disagreement from the point of view that the judges represented two opposed points of view in judging it is interesting to note that disagreement was not considerably larger. The fact that there was 67 per cent agreement indicates the degree of overlap between logic and psychology. This fact is significant because it shows that the above categories are not merely logical but *psychological.*

D. An Objective Free Association Test Scored in Seven Categories

It was estimated above that the irreducible margin of error or disagreement between judges in classifying the free association responses to words is of the order of 20 per cent under the most favorable conditions. One way this margin could be reduced would be to set up an objective test for free association. Terman and Miles (11) used a multiple choice test which they claimed correlated satisfactorily with the ordinary free association test. They presented the stimulus word in the margin followed by four words. The subject was required to check one of the four words.

An objective free association test scored in seven categories was constructed, its reliability was determined and compared with the reliability of the ordinary free association test. This section concerns the construction of the objective test. The data on reliability will be presented in the following section of this article.

Preliminary studies showed that the form of the multiple choice test for free association set up by Terman and Miles (11) contained a significant space error in that there was a strong tendency to check off the first word following the stimulus word. To minimize this error we set the stimulus word above the multiple choices as follows:

<div align="center">

Man.

1. Boy   2. Girl   3. Woman   4. Shoe

</div>

Actually two forms of the multiple choice free association test was developed. The first was the rough draft consisting of 140 stimulus words. The responses were selected on the criteria that they are commonly given for the particular stimulus word. Words having frequencies of over 40 per cent were not used. Although the response words represented seven categories only four response words were given with each stimulus word in order to minimize deliberation on the part of the subject. The response words were so distributed that each category was in the same column the same number of times and the order of categories was equally represented in the test. Test results were obtained from 100 college students with instruction for the test adapted from Terman and Miles. This test was developed by Mr. Sargeant under the supervision of Dr. Henry Odbert (10).

Since the choice of the word responses in this test was quite arbitrary Mr. David Cooper, also working with Dr. Henry Odbert, did an item analysis of Sargeant's data in order to produce a more refined objective test for free association (2). The 100 students who took the Sargeant test were

divided into two groups of 50 for each category. One group represented scores from the upper 50 per cent; the other scores from the lower 50 per cent. The response words were then correlated with the entire test by means of the correlation chart developed by Mosier and McQuitty (7). This was done for each of the categories. The criteria for selection of the final test items were as follows: (*a*) Each of the four responses must be selected by at least 10 per cent of the subjects. (*b*) Each of the responses must correlate with the scale or category of which it is a part, and it should correlate more highly with its respective category than it does positively or negatively with any other scale. This latter requirement may be treated rather roughly inasmuch as differences between correlations are not greatly significant with a population of 100. When an item is not adequate, it might in some cases prove adequate by merely shuffling the scoring of one of the existing items. For example, soldier—sailor, which was used in the test as a contingent contrast, correlates so highly with essential contrast, that it could be retained as an essential contrast, providing none of the other three responses were scored as essential contrast and providing it satisfied with other criteria.

It was hoped that a rigidly selected test of 60 items would be obtained. A test of 59 items was obtained even though some of the items contained only three responses which were valid, the fourth not so. It was decided to use some of these items, scoring only the three valid responses and retaining the fourth merely as a dud to preserve the original pattern of the response words. In the test of 59 items 17 contained a dud word.

The Cooper version of the Sargeant test is refined enough to compare for reliability with ordinary free association test when the same test items are used. The mean of the correlations of the response words with the category for which it stands were above .30 for all categories. The means were considerably higher for essential similarity and essential contrast. The lowest mean correlation was for the operational category which had the smallest number of items because the test was primarily standardized for the six basic categories. The categories were fairly well represented by the response words. All had at least 33 except operationalism with 18. Essential similarity and essential contrast had the largest number of representations, or 37.

E. The Reliability of Free Association and Multiple Choice Techniques Compared

For purposes of comparing the reliability of the multiple choice and the ordinary free association techniques data obtained with the 140 item Sar-

geant test, the 59 item refined Cooper version of the Sargeant test and free associations for the 59 items selected by Cooper are available. The Sargeant test was given to 100 subjects and the reliability was determined by the split half method and the correlations for the various categories were corrected by the Spearman Brown formula (3) in order to derive reliabilities based on the 140 items. The Pearson product-moment formula was used in obtaining correlations. The Cooper version was given to college students and repeated 40 days later. Another group was given an ordinary free association test using the same stimulus words as the Cooper test and also repeated 40 days later. The original intention was to have a population of 100 for all three tests but during the interval of 40 days so many students left for war duties that we were able to obtain a population of only 50 for each test. In order to have the reliabilities comparable the correlations were all corrected for the level of 140 items, thus making all three tests equivalent in terms of number of items. B-K classified the responses for the free association test and they maintained a strictly logical frame of reference.

Table 2 gives the correlation figures for reliability of the seven categories

TABLE 2
RELIABILITY OF CATEGORY CLASSIFICATION

Category	Sargeant Multiple Choice Test	Cooper Multiple Choice Test	Free Association Test
1. Essential similarity	.90	.79	.84
2. General similarity	.55	.65	.81
3. Specific similarity	.73	.84	.88
4. Contingent similarity	.51	.79	.81
5. Essential contrast	.77	.91	.92
6. Contingent contrast	.64	.31	.75
7. Operational	.66	.71	.61
1, 2, 3, 4 grouped	—	.81	.92
5, 6 grouped	—	.91	.92

*All coefficients are corrected for a test of 140 items by the Spearman-Brown formula.

in the three different tests. All correlations are significant except for contingent contrast in the Cooper test. Reliability coefficients for grouped similarities and contrasts are indicated for the Cooper and free association tests.

The correlation coefficients show that the free association technique is more reliable than multiple choice versions. As would be expected the refined Cooper test is more reliable than the rough Sargeant test. There are a few glaring exceptions. The coefficient of .90 for essential similarity on the Sargeant test probably means that as a result of restraints set up

by the criteria which Cooper used in selecting items necessitated the elimination of items which were particularly good for this category. The low coefficient for the contingent contrast category in the Cooper test is out of line though this category has low reliabilities in all three tests. It will be remembered that this category was particularly difficult in the study on the agreement between judges. A possible explanation is that the subject intends to give an essentially contrast response. Sometimes he gives a good contrast and at other times various approximations for essential opposition. However, it is not a matter of chance whether or not the response is essentially opposite or merely in contingent contrast because the correlations are all positive.

Since the populations used in this study are quite small one hesitates to place too much credence in the above correlations but it seems reasonable to conclude that the free association technique is more reliable than the multiple choice technique in spite of errors to which classifiers of free association responses are subject.

F. INTER-CORRELATIONS BETWEEN CATEGORIES

In Table 3 are presented the correlations between the six major categories. The table is in two parts. One section gives the inter-correlations for the Cooper Multiple choice test; the other section for the straight free association test. Since the same stimulus words were used in both tests the correlations are roughly comparable. However, since only 50 subjects were available for both tests the individual correlation figures cannot be taken too seriously. But pattern tendencies in the correlations should be indicative of the direction of the relations. The classification in the free association test was made with the strictly logical frame of responses.

The tables are cut up by means of lines to separate for convenience of inspection of self correlations between the categories, the correlations between the similarity categories, and the contrast categories. The self correrations are along the diagonal line. The inter-correlations between similarity categories are found inside the triangle in the upper half of the table.

The inter-correlations should indicate how selective the categories really are. The self correlations should be higher than the correlations of that category with any other. This requirement is confirmed except in the case of the category of contingent contrast. On the free association test the correlation between essential contrast and contingent contrast is actually higher than the self correlation for contingent contrast. The figures are .62 and .56 respectively. On the Cooper test the figures are .07 and .15. These

TABLE 3

INTERCORRELATIONS BETWEEN CATEGORIES

Cooper Multiple Choice Intercorrelations

	1	2	3	4	5	6
1	+.62					
2	+.11	+.44				
3	—.12	—.13	+.69			
4	—.65	—.09	+.15	+.63		
5	—.67	—.40	—.38	—.06	+.81	
6	+.11	—.13	—.53	—.53	+.07	+.15

Intercorrelations Free Association

	1	2	3	4	5	6
1	+.72					
2	+.38	+.64				
3	—.21	+.13	+.76			
4	—.53	—.12	+.51	+.64		
5	—.20	—.38	—.62	—.56	+.83	
6	—.17	—.66	—.43	—.33	+.62	+.56

figures show that in both forms of the test the contingent contrast category is the poorest.

The fact that the inter-correlation between contingent and essential contrast in the free association test is higher than the self correlation in the contingent contrast category means that these categories are not distinct enough, that there is confusion in classification or that the subject's approximations to contrast overlap these two forms of opposition.

A striking result is the fact that the inter-correlations between either of the contrast categories and any of the similarity categories are negative, except in one instance—between essentially similarity and contingent contrast. This fact indicates that the contrast categories and the similarity categories are distinctive and probably represent different mental processes. Combined with the fact that the reliability figures for grouped similarity and grouped contrast are very high, the indication is that there is a tendency on the part of individuals to respond either in contrasts or in similarities.

The inter-correlations within the similarity categories vary from positive to minus relations. The minus correlations increase in number as we look down the table from Column 1 to 4. It would seem that our arrangement of categories from one to four is in some definite order. Thus categories 1 and 4 are negatively related, also 2 and 4. Adjacent categories tend to be positively related. It would seem that statistically the order from 1 to 4 is an order of decreasing similarity or decreasing levels of abstraction. Another way of stating the problem is to say that the four categories of identification imply an over-lapping hierarchy. Unfortunately the data are not ample enough to determine precisely the order of the hierarchy statistically. This is a suggestive problem and it raises the question of genesis of such a hierarchy.

G. Theoretical Summary

This study indicates that the ordinary free association test is more reliable than an objective substitute in spite of the considerable measure of subjectivity in classifying free association responses. The differences in reliability are not great enough to exclude the possibility of developing an objective substitute for free association which would be just as reliable. How valid would such an objective test be? Terman and Miles (11) found that their objective test correlated as highly with the free association test as the latter correlated with itself. Therefore, statistically, it would seem that the objective test measures the same mental processes as the free association test. Although we were not able to pursue this study to the extent of test-

ing the validity of the multiple choice test there is evidence bearing on this point.

If the objective form of the free association test measures the mental processes of free association then we should expect the number of responses in various categories of the two tests to correlate. The facts show serious discrepancies. Table 4 shows the percentage of the total number of responses elicted by each of the six categories in the two tests. In the multiple choice test a large proportion of the responses are essential identifications; essential contrast elicits the smallest number of choices. In the free association test the contingent identifications are most frequently given. Essential contrast is also popular. The percentages are based on 5811 responses for the free association test and 5259 for the Cooper Multiple Choice test. The difference in the number of responses is due to the fact that the multiple choice test contained 17 duds.

TABLE 4

PERCENTAGE OF THE RESPONSES GIVEN FOR EACH CATEGORY FOR THE FREE ASSOCIATION AND MULTIPLE CHOICE TESTS

Categories	1	2	3	4	5	6
Free asso. test	.158	.126	.168	.296	.187	.065
Multi. choice test	.400	.147	.167	.121	.071	.093

The above discrepancies fit the following explanation. In the multiple choice test the subject's determining tendency is to match the response word with the stimulus word and in most instances the most plausible match is the response which is essentially similar to the stimulus word. The less similar the response the smaller number of choices; the least similar response is the essential contrast type and this category has the smallest number of elections. When the objective test was given it was noticeable that the subjects deliberated longer than they ordinarily do when free associating. This tendency to match the response with the stimulus word is also present in free association but the process here is more spontaneous and contingent associations play a larger part. Moreover, association by contrast is known to be frequent in free responses to words. The objective test seems to inhibit this tendency. In order to develop an objective free association test some system of weights, or control of the opportunities to choose each category, must be applied to obtain results more closely representing the categorical distribution of the free association test.

From the point of view of the kind of mental process that is involved in association the fact that the contingent contrast category has low relia-

bility for both the objective and subjective forms of association is important. There seems to be an overlap between this category and essential contrast for both the judges who classified the associations and the subjects who gave them. Logically this category is just as valid as contingent identification and yet the latter stands up statistically as a reliable category. The difference may be that in ordinary thinking contrasting implications are relatively vague and undifferentiated. Logically we might have paralleled the identifications with four contrasts. Specific contrast is a possible category as is indicated by such an example such as white–coal. However, it never occurred to the writers to set up such a category and it is unlikely that many responses of this type occur. A counterpart of general identification such as general contrast apparently does not exist in our thinking although it is logically predicated. Associated thinking is oriented in a positive direction of sensing analogies at various levels of identification. Negative associations tend to be sensed merely as oppositions without clear-cut differentiations of the aspect to which the oppositions is made. From the point of view of semantics the interesting question is raised as to the implications for thinking and action if our use of negative implications were as fully developed as positive or analogous relations.

REFERENCES

1. BRIDGMAN, P. B. The Logic of Modern Physics. New York: Macmillan, 1937. Pp. 228.
2. COOPER, D. The refinement and construction of a word association test to be objectively scored in six categories. Major thesis, Department of Psychology, Dartmouth College, 1942.
3. GARRETT, H. E. Statistics in Psychology and Education. New York: Longmans, Green, 1940. Pp. 493.
4. KARWOSKI, T. F., GRAMLICH, F. W., & ARNOTT, P. Psychological studies in semantics: I. Free association reactions to words, drawings, and objects. *J. Gen. Psychol.*, 1944, **20**, 233-247.
5. KARWOSKI, T. F., & ODBERT, H. S. Color-music. *Psychol. Monog.*, 1938, **50**, No. 2, 1-60.
6. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. F. Studies in synaesthetic thinking: II. The rôle of form in visual responses to music. *J. Gen. Psychol.*, 1942, **26**, 199-222.
7. MOSIER, C., & McQUITTY, J. V. Methods of item validation and abc's for item-test correlation and critical ration of upper-lower difference. *Psychometrica*, 1940, **5**, 57-65.
8. ODBERT, H. S., KARWOSKI, T. F., & ECKERSON, A. B. Studies in Synaesthetic thinking: I. Musical and verbal associations of color and mood. *J. Gen. Psychol.*, 1942, **26**, 153-173.
9. PLATO. Dialogues (trans. by Jowett). New York: Scribner, 1907. Pp. 433-435.

10. SARGEANT, D.   Word association. Major thesis, Department of Psychology, Dart-
     mouth College, 1942.
11. TERMAN, L. M., & MILES, C. C.   Sex and personality. New York: McGraw-Hill,
     1936. Pp. 600.

Department of Psychology
Dartmouth College
Hanover, New Hampshire